Monitoring Pharmacy Expert System Performance Using Statistical Process Control Methodology

Joshua A. Doherty, B.S. [†+], Richard M. Reichley, R.Ph.[† +], Laura A. Noirot, B.S. [†+], Ervina Resetar, M.S. [†+], Michael R. Hodge, M.A.[+], Robert D. Sutter, RN, MBA, MHA,[+] Wm Claiborne Dunagan, M.D. [†+], and Thomas C. Bailey, M.D. [†+]

Medical Informatics Laboratory, Washington University School of Medicine[†] and the Center for Quality Healthcare and Effectiveness[+], BJC Healthcare, St. Louis, Missouri

ABSTRACT

Automated expert systems provide a reliable and effective way to improve patient safety in a hospital environment. Their ability to analyze large amounts of data without fatigue is a decided advantage over clinicians who perform the same tasks. As dependence on expert systems increase and the systems become more complex, it is important to closely monitor their performance. Failure to generate alerts can jeopardize the health and safety of patients, while generating excessive false positives can cause valid alerts to be dismissed as noise. In this study, statistical process control charts were used to monitor an expert system, and the strengths and weaknesses of this technology are presented.

INTRODUCTION

The Medical Informatics Laboratory (MIL) at the Washington University School of Medicine (WUSM) and BJC HealthCare (BJC) has developed several expert systems. As more of these systems come online, it is a challenge to monitor their performance. Finding an effective and efficient means to carry out this function is particularly important given the scope of the newer systems being developed. Early expert systems developed by our group screened drug orders using only a few hundred rules. Our newer pharmacy expert systems use thousands of drug rules, and manually monitoring the performance of each one would be essentially impossible.

One solution to this problem is to use statistical process control (SPC) to automate monitoring of expert system performance. By using statistical analyses to monitor rules, attention can be focused on only those rules that exhibit significant changes in performance over time. This approach can reduce the amount of manual energy that must be spent on performance monitoring.

Originally developed for use in the manufacturing industry, the principles and techniques of SPC can easily be applied to expert systems. SPC methods have also been advocated in healthcare for quality improvement[1] and for healthcare epidemiology.[2,3] Our group has previously used control charts and other SPC tools for manually monitoring expert system performance and impact.[4,5]

METHODS

We selected our DoseChecker application to test whether SPC could be used to automate expert system monitoring. DoseChecker was designed to screen drug orders for dosing errors that result from failing to adjust for renal function.[6] Using patient-specific information from pharmacy and laboratory systems, a creatinine clearance estimate is calculated and the ordered dose is compared to a set of allowable dose ranges. If a dose violates one of the rules, a fax or page notifies the pharmacist responsible for the patient. In addition to the order and patient information, the expert system also provides the pharmacist with a recommended dose appropriate for the patient's age, weight, and renal function.

After addressing an alert, the pharmacist uses a dynamic web interface to enter response information into a database, including whether or not the pharmacist and physician agreed with the alert. This information is necessary to effectively monitor DoseChecker since there are valid clinical reasons why a patient can be given a drug dose that is outside of the recommended ranges.

In evaluating strategies for automating DoseChecker monitoring, our team chose SPC methods as the most efficient and effective solution. The principle tool of SPC is the control chart, which plots performance data over time along with calculated upper and lower control limits.

The control limits define the variation that can be attributed to common causes, and in our case are set at three standard deviations (3σ) above and below the

process average. If a point falls outside the control limits or if other unexpected patterns occur in the data, a change has occurred that cannot be attributed to random fluctuations. This is called special cause variation, and may indicate that a significant change or event has occurred in either the expert system or the clinical environment.

For this application, we employed a p-chart using three rules for signaling special cause variation[7,8]:

1. A point above or below the calculated control limits.
2. Eight points in a row above or below the process average.
3. Any ten out of eleven points above or below the process average.

Any single data point can satisfy the first rule. The second and third rules both require a series of data points, which indicate a sustained change in the process. When this occurs, the process average may be recalculated in order to evaluate future performance.

Figure 1 is a sample control chart for a process that shows no special cause variation. The centerline is the process average, while the dotted lines above and below are the control limits. The control limits vary with each point on the chart, because they are dependent on the sample size used to calculate the individual monthly rates. The points connected by the solid black line represent the actual data used to calculate the other aspects of the control chart. Each point represents a single month.

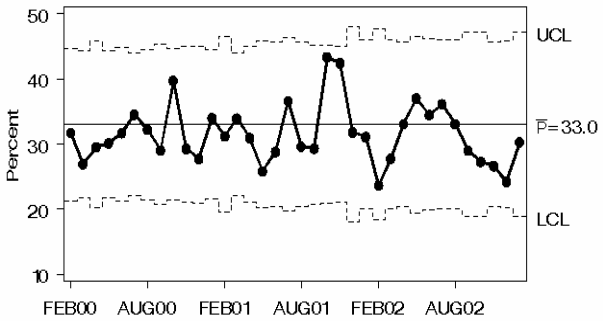

Figure 1: Sample Alert Rate Chart Without Special Cause Variation

For the initial trial, the following DoseChecker performance attributes were selected for monitoring:

- Alert Rate
- Pharmacist Agree Rate
- Physician Agree Rate
- Missing Outcome Response Rate

The Alert Rate is the number of alerts divided by the number of orders screened. The Pharmacist Agree Rate is a subset of the alert rate and is calculated by dividing the number of alerts with which the pharmacist agreed by the number of alerts for which an outcome was entered. The Physician Agree Rate is a smaller subset and is calculated by dividing the number of alerts with which the physician agreed by the number of alerts with which the pharmacist agreed. If the pharmacist does not agree with an alert, a physician is not contacted. Finally, the Missing Outcome Response Rate is the number of alerts for which no outcome was entered divided by the total number of alerts.

To create the control charts, a monthly procedure was put in place to analyze the necessary information, using a standard statistical software package (SAS, Cary, NC). The same process performs the calculations and generates the control charts, which are displayed via an existing intranet web application. To make problems easier to detect, control charts that show special cause variation trigger an SPC flag to appear on the website. This flag is a small icon that appears next to the drug name and lets the user know when any of the three special cause rules have been satisfied. The flags were the basis for our analysis, since they could be used to notify our maintenance team that a specific drug or drug rule should be investigated.

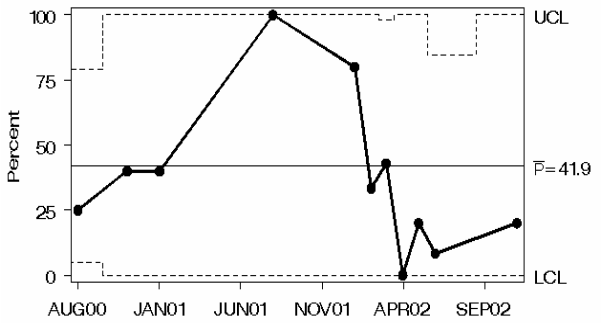

Figure 2: Ticarcillin/Clavulanate Alert Rate

Because of the statistical calculations involved in generating SPC charts, only drug rules with an average of twenty-five or more orders screened each month were evaluated. Rules with very small sample sizes tend to have highly variable data, which makes the control limits on the SPC charts much less useful. Figure 2 shows a control chart for the Ticarcillin/Clavulanate Alert Rate at one BJC HealthCare facility. With a twelve-month average of fewer than four orders per month, the result is a chart with control limits that range from zero to one hundred percent.

RESULTS

As of September 2002, automated SPC monitoring of DoseChecker was operational for all five of the BJC facilities where DoseChecker was deployed. Our approach was to take problems that we found manually over the last year and determine whether a routine analysis of the SPC flags would have helped us discover them. We accomplished this by examining control charts generated from historical data prior to September 2002. During the course of our analysis, we were able to find several examples where control charts would have been very effective in identifying unusual behavior in the expert system. However, we were also able to identity a number of issues that would need to be addressed before such a system could be used to routinely automate performance monitoring.

Figure 3 is one example of how SPC methods can be used to detect a process that is out of statistical control. In this case, a staffing change at one BJC facility in November 2001 prevented their pharmacy from effectively addressing DoseChecker alerts and entering response information. This control chart for intravenous vancomycin was generated the month following the staffing change. The Missing Outcome Response Rate increased to forty percent, which is nearly ten times the value of the upper control limit. This triggered an SPC flag that could have been used to alert the maintenance team of a potential problem.

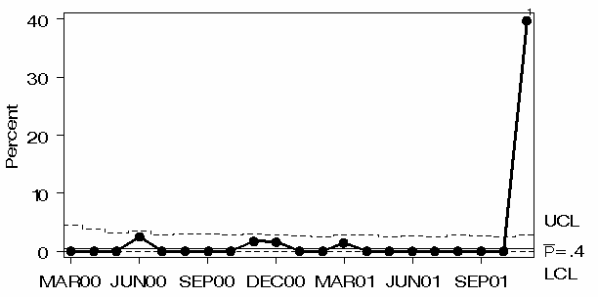

Figure 3: Missing Outcome Response Rate

Another example that illustrates how SPC can be used for monitoring an expert system occurred early in 2002, when policy changes at a BJC facility gave the pharmacists the authority to change certain drug orders in response to DoseChecker alerts. Prior to this, the Pharmacist Agree Rate for intravenous vancomycin alerts was nearly zero. Afterwards, the Agree Rate increased dramatically, as seen in Figure 4. This increase generated another SPC flag that would have notified the maintenance team.

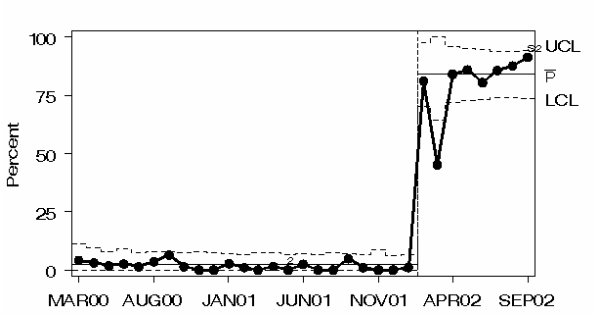

Figure 4: Increase in Pharmacist Agree Rate as a Result of a Policy Change

This control chart also shows the benefit of recalculating a shift in the process average. After eight points were plotted that were above the previous average, a new average was calculated using the same eight points. If the new process average were not calculated, the automated monitoring system would continue to generate SPC flags until the agree rate fell to its previous rate. In this case, the policy change resulted in an improved process because pharmacists were agreeing with the alerts, so expecting the process to return to the previous state would not make sense. Furthermore, by redefining the process average, an SPC flag would be generated if the agree rate decreased below its newly established baseline.

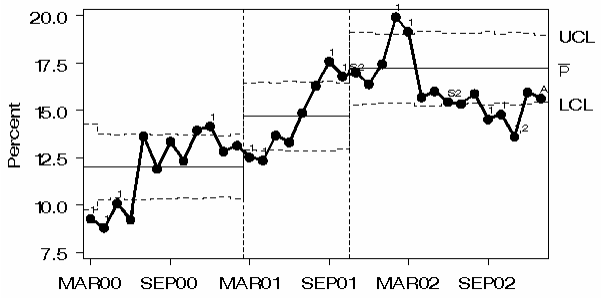

Figure 5: Change in Overall Alert Rate

Finally, we saw that control charts can draw attention to overall trends at a facility. Figure 5 shows the overall Alert Rate at a BJC facility as of January 2003. From March 2000 to March 2002, three different process average shifts occurred. Each shift resulted in an SPC flag that could have called attention to the change in the alert rate. After the facility implemented a policy change early in 2002, three points have been below the lower control limit, and eight in a row and ten out eleven have been below the mean. All of these events result in SPC flags that indicate a statistically significant trend in

response to the policy change.

DISCUSSION

The use of SPC tools to monitor and evaluate expert systems is not a new concept to our group. We have used control charts to monitor nurse disagreement rates with our GermWatcher application, to help ensure that constant modifications to the system did not result in a deterioration in performance.[4] We have also used SPC methods to show the impact of our PharmADE application in reducing the number of cisapride drug interactions at our pilot hospital.[5]

What makes this project different from earlier work is the implementation of an automated system capable of monitoring a large rule set and notifying a maintenance team when a process showed significant variation from expected performance. If successful, such an application could be scaled to handle even larger rule sets, resulting in increased efficiency and lower maintenance costs. By focusing our study on events that we knew had impacted DoseChecker, we were able to provide evidence that such an automated monitoring system could be effective. These events were all significant, and each could have dramatically affected the performance of DoseChecker if they had gone undetected.

For example, the increase in the Missing Outcome Response Rate in November 2002 had important consequences, since an expert system is ineffective if alerts are not being addressed. While such drastic occurrences are not common, this system could detect smaller changes that might occur if a pharmacist began entering less response information because of time constraints or other factors.

The rising alert rate seen in Figure 5 is another example of an event that could cause problems if undetected. Discovering a trend such as this is important because a high alert rate can have a negative affect on performance due to "alert fatigue."[9] Prior work suggests that a poor signal to noise ratio can greatly limit the utility of automated alerts.[10]

Our automated monitoring system would have alerted the maintenance team to both of these problems, in addition to the positive Agree Rate change seen in Figure 4. Although we were unable to find examples, SPC methods can also find potential problems with an expert system that might not be detected by manual review of raw data. When data are highly variable, manually detecting statistically significant changes can be very difficult. Using control charts allows problems to be detected before they reach a point at which they are obvious to both the users and the maintenance team.

While our analysis illustrates the potential benefit of SPC monitoring, we also encountered several important implementation issues. First, while control charts are very good at detecting special cause variation in a process, they do not provide insight into what caused the variation. This is further complicated by the fact that the variation could represent a measurement problem, software performance issue, or an actual process change. In order to effectively understand the variation, there must be well-established communication channels between the software development and maintenance teams and clinical user representatives.

Another issue concerns when to recalculate the process average. A major concern is that if care is not taken to understand the cause of the shift, an underlying problem can be hidden by the suppression of future alerts. For example, if an alert rate rises in a sustained manner, it will eventually reach a point when a statistically significant shift occurs in the process mean. At that point, future alerts will stop unless the rate continues to increase enough to either again recalculate the mean or the new control limits are exceeded.

Analysis of the shift might show that new prescribing practices require a minor change in the expert system rule. By making this change the alert rate would fall back to normal and the mean would eventually shift back down again. However, if the analysis was not performed correctly, the lack of future alerts would make it appear as though the process were functioning optimally when it was not. Therefore, we would suggest that the process average not be automatically recalculated for shifts that suggest degradation in process performance. Instead, such shifts should serve as a signal to investigate the process.

The sampling timeframe is another important issue to consider. With our current application the control charts are calculated monthly, potentially limiting the timeliness of detecting important process changes. Ideally, data could be plotted more frequently so that problems can be resolved in a shorter amount of time. However, important factors that are relevant to the selection of the sampling timeframe include the frequency of events, resource constraints limiting the response to SPC signals, and clinical risk/benefit issues.

Currently, the system is limited to those drugs rules that screen an average of twenty-five or more orders in a month. The remaining drugs are not monitored at all. This is because of problems encountered when working with a small sample size as illustrated in Figure 2. One possible solution would be to change the sampling frame such that each data point on the control chart represents a given number of orders screened rather than a period of time. This would result in control charts being generated more often for high volume drugs, and less frequently for low volume drugs.

A final limitation of our current application is that statistical analysis is not done on the number of orders screened each month. For example, in August of 2002, one facility stopped using gatifloxacin because of a formulary change. This is exactly the same situation that would have occurred had DoseChecker malfunctioned and stopped screening a particular drug. Because the number of alerts fell in proportion to the number of orders screened, no significant change was seen in the Alert Rate or the other monitored statistics. This is relatively simple to correct in comparison to the previous issues discussed.

CONCLUSIONS

SPC Methodology is an efficient and effective means to automate the monitoring of expert system rule behavior and other process measures that are important for expert system performance.

Supported in part by AHRQ Grant U18 HS11898-01

REFERENCES

1. Maleyeff J. Kaminsky FC. Jubinville A. Fenn CA. A guide to using performance measurement systems for continuous improvement. *Journal for Healthcare Quality.2001; 23(4):33-7*
2. Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology, part I: Introduction and basic theory. *Infection Control & Hospital Epidemiology*. 1998; 19(3):194-214.
3. Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology, Part II: Chart use, statistical properties, and research issues. *Infection Control & Hospital Epidemiology*. 1998 19(4):265-83.
4. Kahn MG, Bailey TC, Steib SA, Fraser VJ, Dunagan WC. Statistical process control methods for expert system performance monitoring. *JAMIA 1996;3:258-269*.
5. McMullin ST, Reichley RM, Watson LA, Steib SA, Frisse ME, Bailey TC. Impact of a web-based clinical information system on cisapride drug interactions and patient safety. *Arch Intern Med. 1999;159:2077-2082*
6. McMullin ST, Reichley RM, Kahn MG, Dunagan WC, Bailey TC. Automated system for identifying potential dosing problems at a large university hospital. *Am J Healthcare Pharm Assoc. 1997;54:545-9*.
7. Wheeler, DJ. Advanced topics in statistical process control. Knoxville (TN): SPC Press, Inc; 1995.
8. Grant, EL, and RS Leavenworth. Statistical quality control. 6th ed. New York (NY): McGraw Hill; 1980.
9. Peterson JF, Bates DW. Preventable medication errors: identifying and eliminating serious drug interactions. *J Am Pharm. 2001; 41:2:156-160*.
10. Glassman PA, Simon B, Belperio P, Lanto A. Improving recognition of drug interactions: benefits and barriers to using automated drug alerts. *Med Care. 2002; 40:12:1161-1171*.